Traditional Clustering with SCSI (non-IOP)
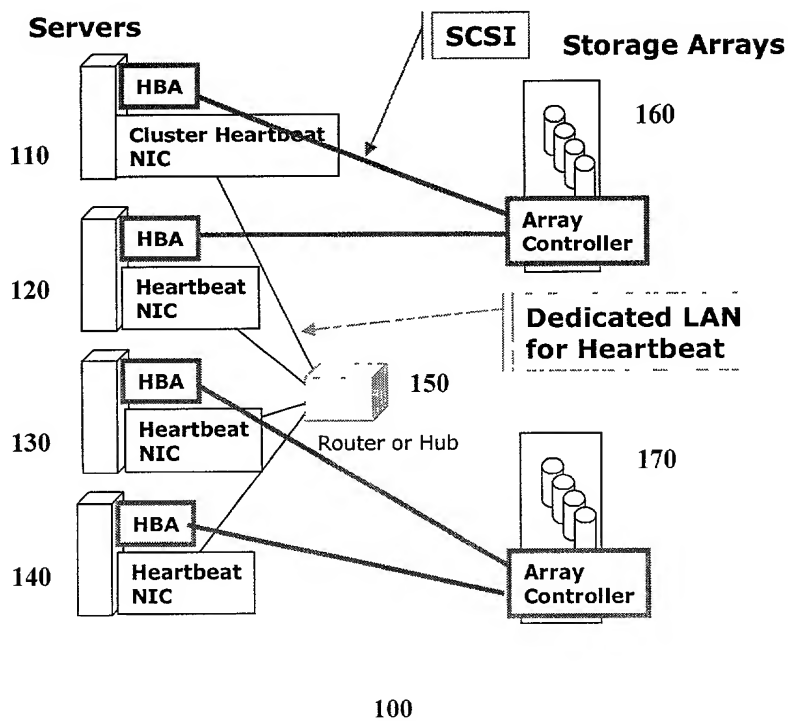
Servers   **SCSI**   **Storage Arrays**

HBA

Cluster Heartbeat NIC

110

160

HBA

Heartbeat NIC

120

Array Controller

Dedicated LAN for Heartbeat

HBA

Heartbeat NIC

130

150

Router or Hub

170

HBA

Heartbeat NIC

140

Array Controller

100

FIG. 1

PRIOR ART

Fibre Channel based Clustering Topology

Servers

Storage Arrays

HBA

Cluster Heartbeat NIC

210

FC

200

270

Array Controller

HBA

Heartbeat NIC

220

FC Switch

250

280

Array Controller

HBA

Heartbeat NIC

230

Router or Hub

260

290

Dedicated LAN for Heartbeat

HBA

Heartbeat NIC

240

Array Controller

FIG. 2

PRIOR ART

IOP based Clustering Topology

Servers

IOP monitors Local Server Heartbeat

Storage Arrays

310 IOP

360

Gig-Ethernet iSCSI

IOP

320 IOP

370

330 IOP

Storage Router

IOP — IOP

350

IOP

380

340 IOP

300

IOP

HEALTH MONITORING AND HEARTBEAT LOGIC CIRCUIT 392	FAILURE/RECOVERY LOGIC CIRCUIT 394
CLUSTER NODE ADD / REMOVE LOGIC CIRCUIT 396	CLUSTER MEMBERSHIP DISC./RECONC. CIRCUIT 398

INPUT/OUTPUT PROCESSOR (IOP)

FIG. 3
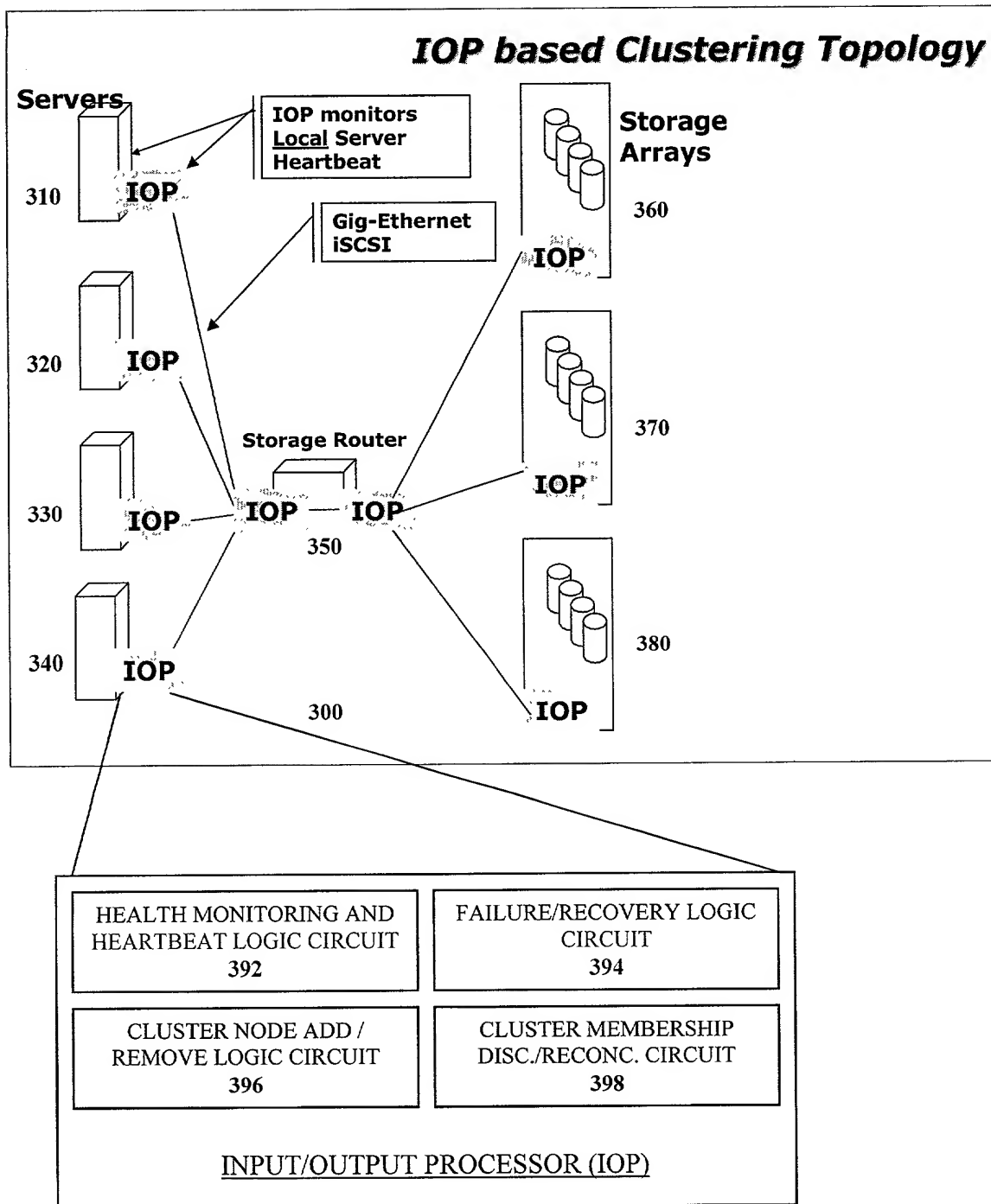
Cluster Failure/Recovery Logic
(High level representation)

Start of Day
    A-Active   410
    B-Active
    C-Active
    D-Standby

Node A Fails
    A-Down ←   420
    B-Active
    C-Active
    D-Standby

Node D Takes over (mounts storage, starts EXE, assume floating IP addr)
    A-Down
    B-Active   430
    C-Active
    D-Active ←

Node A Recovers
    A-Standby ← 440
    B-Active
    C-Active
    D-Active

Server Farm Functionality:
After "A" recovers, it
becomes the standby

FIG. 4

Cluster Heartbeat and Health Monitoring (system level representation)

Server A	**Server B**	**Server C (stand-by)**	**Storage X**	**Storage Y**

Time

510
- A-Local IOP checks System response
- A-Local IOP produces A-OK

| B-Local IOP Checks system response | C-Local IOP Checks system response | Local IOP Checks system response | Local IOP Checks system response |

No heartbeat required until a problem arises

520
- A-Local IOP Checks System Response Gets NO response
- A-Local IOP produces DOWN

- B-Local IOP receives A-DOWN
- Update ClusterMIB

- C-Local IOP receives A-DOWN
- Update ClusterMIB Failover Mount storage Start Executables Assume IP alias
- C-Local IOP produces C-OK

530

540
- B-Local IOP receives C-OK
- Update ClusterMIB

- X-Local IOP receives C-OK
- Update ClusterMIB

- Y-Local IOP receives C-OK
- Update ClusterMIB

550
- A Server Recovers A-Local IOP Is aware that A is healthy
- A-Local IOP produces A-stand-by

- B-Local IOP receives A-stand-by
- Update ClusterMIB

- C-Local IOP receives A-stand-by
- Update ClusterMIB

- Local IOP receives A-stand-by
- Update ClusterMIB

- Local IOP receives A-stand-by
- Update ClusterMIB

Failover complete
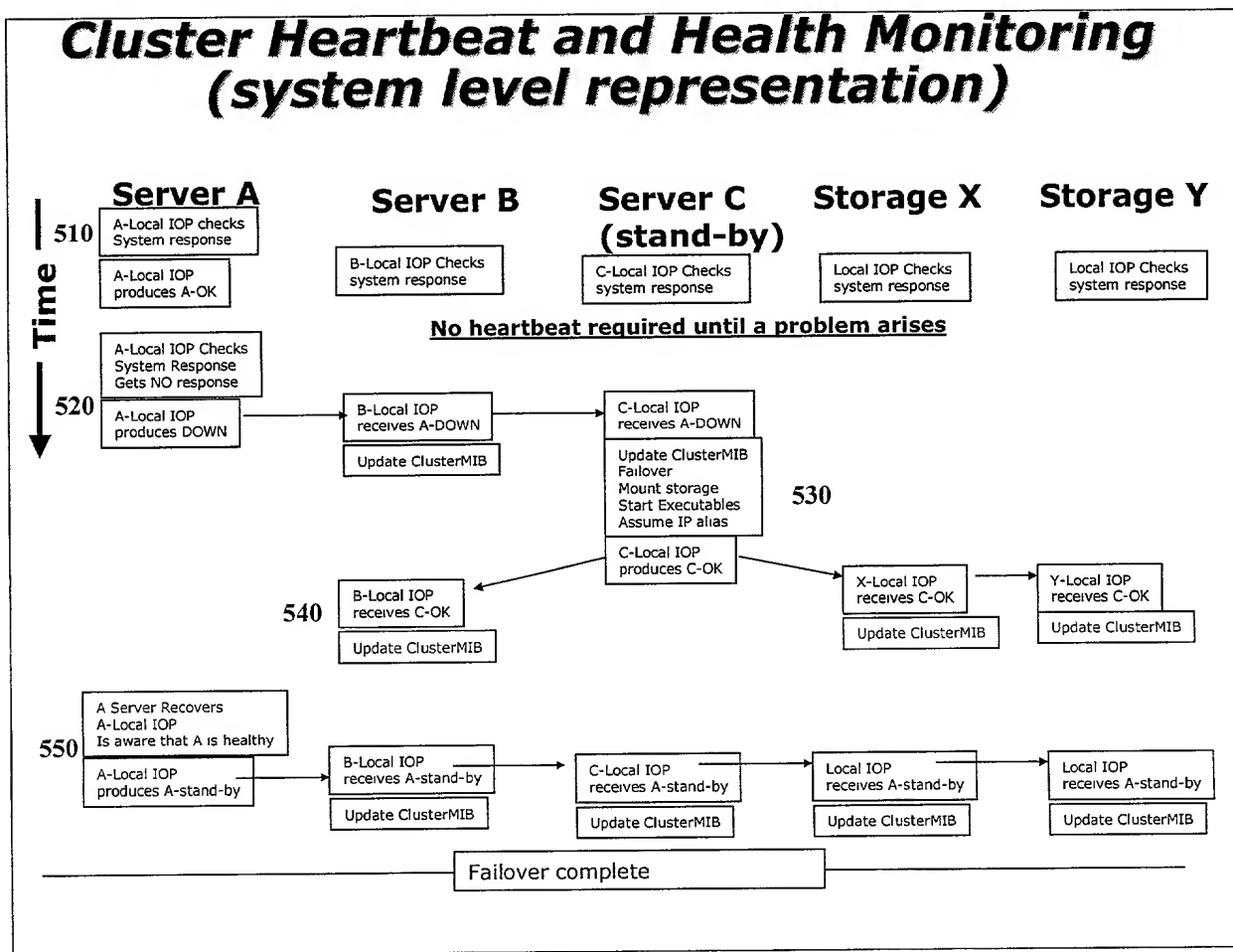
FIG. 5

Cluster Node add/remove
(High level representation)

Start of Day
    A-Active   610
    B-Active
    C-Active
    D-Standby

Add Node E   620
    A-Active
    B-Active
    C-Active
    D-Standby
    E-Down   ⟵

Confirm E will work (test mount storage, start EXE, check floating IP addr)
    A-Active
    B-Active   630
    C-Active
    D-Standby
    E-Standby ⟵

Put Node D to work (optionally run with 2 standbys)
    A-Active
    B-Active   640
    C-Active
    D-Active ⟵
    E-Standby

> **Server Farm Functionality:**
>   Add/Remove node
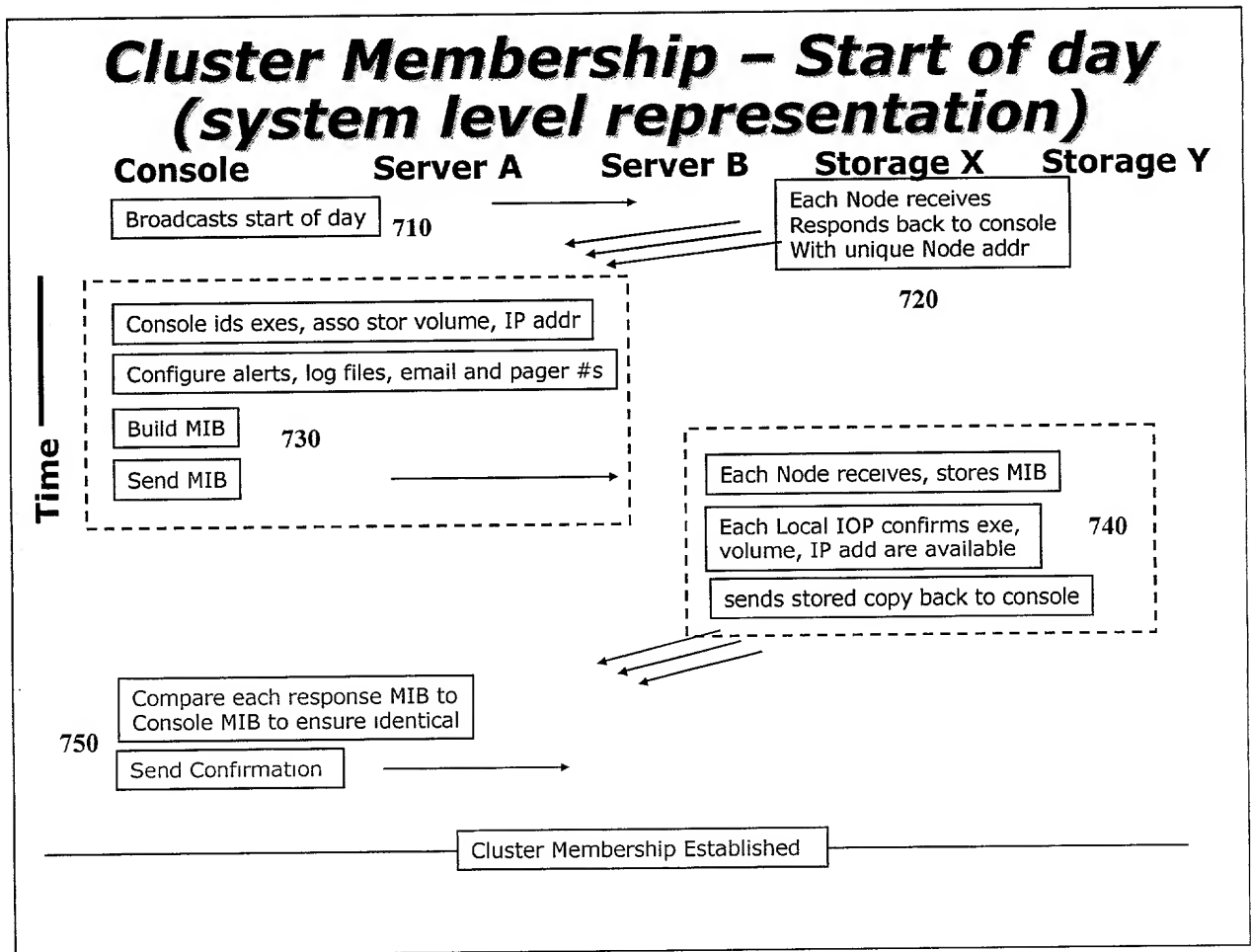>   without taking cluster off
>   line

FIG. 6

Cluster Membership – Start of day (system level representation)

Console	Server A	Server B	Storage X	Storage Y

Broadcasts start of day   710

Each Node receives
Responds back to console
With unique Node addr

720

Time

Console ids exes, asso stor volume, IP addr

Configure alerts, log files, email and pager #s

Build MIB   730

Send MIB

Each Node receives, stores MIB

Each Local IOP confirms exe,
volume, IP add are available   740

sends stored copy back to console

Compare each response MIB to
Console MIB to ensure identical

750   Send Confirmation

Cluster Membership Established

FIG. 7